

Statistical Multiplexing of H.264 Video Streams Using Structural Similarity Information^{*}

LUÍS TEIXEIRA^{1,2} AND LUÍS CORTE-REAL²

¹*Research Center for Science and Technology in the Arts*

Universidade Católica Portuguesa

Porto, 4169-005 Portugal

²*Faculdade de Engenharia/INESC Porto*

Universidade do Porto

Porto, 4200-465 Portugal

In this paper, we propose a method to broadcast digital video programs in which the channel capacity is dynamically distributed among video programs according to each video program particular complexity. A bit rate control algorithm based on the Structural Similarity Index as the measure of video program complexity is examined. Initial results show that a uniform picture quality among video programs can be obtained.

Keywords: subjective quality, structural information, joint video coding, H.264, statistical multiplexing

1. INTRODUCTION

Several international video coding standards such as MPEG-1, MPEG-2, MPEG-4, H.263 and H.264 [1] have been developed in the last years. These standards, based on hybrid motion-compensated predictive video coding, have improved coding efficiency in applications from broadcast to storage.

The H.264/MPEG-4 Advanced Video Coding standard (H.264/AVC) [1], also referred as ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 Part 10), is the latest video coding standard jointly developed by the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). H.264/AVC has accomplished a considerable progress regarding coding efficiency, error robustness, and increased flexibility and broader scope of applicability relative to its predecessors [2]. It covers all frequent video applications ranging from mobile services and videoconferencing to IPTV, HDTV, and HD video storage [2]. In TV multi-program broadcast systems, the use of H.264/AVC standards allows bandwidth economies during the transmission of programs, while providing a service with a higher quality regarding current systems [3].

In a typical TV broadcast service, such as multiple live (news, sports or live shows) or pre-encoded (*e.g.* movies, advertisements, *etc.*) video programs are multiplexed onto a single constant bit rate channel. Initial implementations used a straightforward approach: existing capacity of broadcasting channel was equally divided among all programs, and

Received February 1, 2008; accepted November 28, 2008.

Communicated by Yau-Hwang Kuo, Pau-Choo Chung and Jar-Ferr Yang.

^{*}This work has been supported by “Fundação para a Ciência e Tecnologia” and “Programa Operacional Ciência e Inovação 2010” (POCI 2010), co-funded by the Portuguese Government and European Union by FEDER Program.

each sequence was coded independently at constant bit rate (CBR). In CBR, the quality of the video sequence varies due to the variations in the scene complexity.

One possible solution is to allocate different bit rates to each video encoder based on the expected image complexity of the signal to be encoded. Such an allocation scheme should be dynamically adjusted over time, depending on the relative complexity of each channel. This process is called statistical multiplexing (stat-mux).

In this paper, we propose a statistical multiplexing rate control scheme for H.264 coding of video sequences. To obtain uniform picture quality within multiple video programs, a novel complexity measure based on Structural Similarity Information is proposed, which can assess the coding complexity of the current frame regarding its perceptual quality. Furthermore, it can be implemented in each video encoder in a very simple way without resulting in an excessive increase in the computational power.

In section 2, Video Quality Assessment is introduced regarding objective and subjective video quality. The approaches are presented: Structural Similarity Index (SSIM) and Subjective Assessment Methodology for Video Quality (SAMVIQ). In section 3, the proposed scheme is described and experimental results presented.

2. VIDEO QUALITY ASSESSMENT

Video programs are produced to be displayed to human viewers. As a result it is their opinion regarding the video quality that is relevant. Informal and formal subjective measurements have always been, and will continue be used to evaluate system performance from the design lab to the operational environment [4, 5]. Although the need for an objective testing method of picture quality is clear, subjective testing is too complex and provides too much variability in results. However, due to the importance of the observer's opinion of picture quality, any objective measurement system must have good correlation with subjective results for the same video system and test scenes. To be able to incorporate Human Visual System (HVS) model into broadcasting encoding system can result in additional improve of the coding efficiency and enhance video quality.

PSNR and MSE are amongst the most popular employed objective quality metrics due to their low complexity and clear physical meaning. Nevertheless, both metrics have been target of an high number of critics for not correlating well with HVS [6, 7]. This is due to the fact that they can not signify the exact perceptual quality as they are based on pixel to pixel difference calculation and ignore human perception and viewing condition. The impact of coding distortion on the subjective quality is still under investigation [8-10]. A new scheme for a class of quality metrics, known as: Structural Similarity Index (SSIM) has been proposed to models perception implicitly by taking into account the fact that the HVS is adapted for extracting structural information (relative spatial covariance) from images [6].

2.1 Structural Similarity Index (SSIM)

SSIM is an objective image quality assessment metric which attributes perceptual degradations to structural distortions [11]. The SSIM index has been demonstrated in [5] to be an effective measurement of perceptual global degradations in natural images.

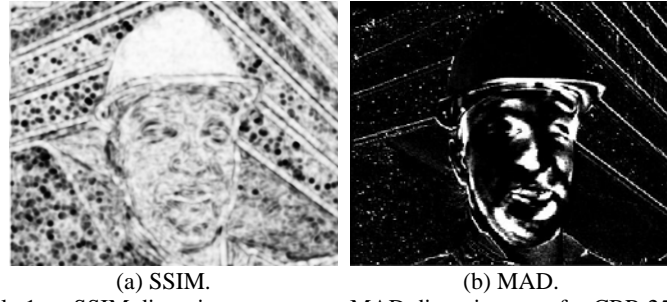


Fig. 1. Example 1 on SSIM distortion map versus MAD distortion map for CBR 256kbps, frame 1, foreman sequence.

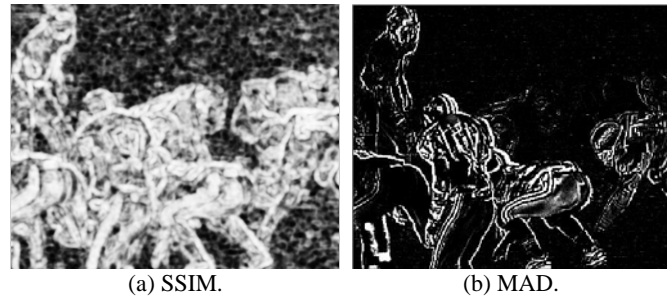


Fig. 2. Example 2 on SSIM distortion map versus MAD distortion map for CBR 256kbps, frame 1, football sequence.

Figs. 1 and 2 compare the use of SSIM metric with maximum absolute distortion (MAD) metric. In both cases, SSIM provides local perceptual cues of importance that can support towards perceptual visual coding. Brighter regions are associated with regions that present a better visual quality. The SSIM Distortion Map clearly captures the loss of quality in the grass and in the football player's clothes, *etc.* The MAD Error Map does not represent the distortion existing in the different regions of the image. This is a result of SSIM has successfully incorporates HVS characteristics without much added complexity. The local SSIM index measures the similarities of the local luminance $l(x, y)$, contrast $c(x, y)$, and structure $s(x, y)$ using the following equations:

$$l(x, y) = \left(\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \right), c(x, y) = \left(\frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \right), s(x, y) = \left(\frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \right) \quad (1)$$

where x and y are two nonnegative image signals to be compared, μ_x and μ_y are the mean intensity of image x and y correspondingly, σ_x and σ_y are the standard deviation of image x and y respectively, σ_{xy} is the covariance of image x and y , C_1 , C_2 and C_3 are small constants to avoid instability when $\mu_x^2 + \mu_y^2$ are very close to zero [11, 12]. The SSIM measure is thus a combination of these three distortion components between signals x and y as shown below:

$$SSIM(x, y) = l(x, y)^\alpha \cdot c(x, y)^\beta \cdot s(x, y)^\gamma \quad (2)$$

where α , β and γ are positive parameters and are used to regulate the relative importance of each of the components. Typically these values are set to value one [12] and

$$C_1 = (K_1 L)^2, C_2 = (K_2 L)^2, C_3 = C_2/2 \quad (3)$$

where $K_1, K_2 \ll 1$ and L is the dynamic range of the pixel values (for 8-bit grayscale image this value is 255) [12]. SSIM index can thus be computed by the following equation:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (4)$$

The index is obtained for all the pixels in the image resulting in a quality map. The final step is to convert the quality map into a single quality index. This can be achieved by computing the mean value of local SSIM values or by other pooling methods [13, 14]. If we consider $SSIM(x_i, y_i)$ the local index value at the i th local sample then

$$SSIM(X, Y) = \frac{\sum_{i=1}^{N_s} W_i(x_i, y_i) \cdot SSIM(x_i, y_i)}{\sum_{i=1}^{N_s} W_i(x_i, y_i)} \quad (5)$$

where N_s is the number of samples in the quality map, and $W_i(x_i, y_i)$ is the weight given to the i th sample. In our case the weight was constant so we have obtained the mean SSIM (MSSIM) index [5].

2.2 Subjective Assessment Methodology for Video Quality (SAMVIQ)

In order to assess the visual part of multimedia codecs or systems, EBU Project Group B/VIM has proposed a new subjective evaluation methodology called SAMVIQ (Subjective Assessment Methodology for Video Quality). The SAMVIQ method have been standardized at EBU and successfully tested by several labs involved in BVIM (Broadcast Video In Multimedia) study group the in the last years [15, 16].

SAMVIQ is a multi stimuli continuous quality scale method using explicit and hidden references [15, 16]. It generates a measure of the subjective video codec quality which can be compared directly with the reference, *i.e.* the original video signal. Test sessions are organized in such a way that each scene is rate in a sequential order. For each scene, participants have to assess various algorithms including an explicit reference and a hidden reference. The participant can start or stop the evaluation, give, alter or maintain the current score for each video clip when he wants. Furthermore, he can view sequences as many times as he wants. A participant can only give a mark after viewing the complete clip. After the participants have assessed all the clips of one scene he can carry with the next scene of the test evaluation. In our case, in each session of SAMVIQ 2 codecs were used (our approach and JM Reference Model) resulting in the participant able

to select and view among 5 clips: 2 with the impaired signals, and the remaining 3 consisting in the known reference, the “hidden” reference and the “hidden low” anchor.

Testing took place during September 2007. In total 15 evaluators, aged 20-61 participated in the test. All participants had normal or corrected-to-normal visual acuity. Four “experimented” viewers participated in the evaluation. These “experimented” viewers have experience, knowledge, motivation and patience to consider the test sequences viewed in a critical way. This follows EBU SAMVIQ recommendation that at least half of subjects should be “non-expert” viewers [15, 16]. Two bit rates were evaluated. The low anchor was encoded with H.264 codec with fixed quantization step-size of 42. Explicit reference and hidden reference correspond to the same file: the original file. Video clip was displayed in normal resolution, on a 17” CRT display at a viewing distance of around 5H (H is the picture height), and the computer with a resolution of 1024×768 pixels.

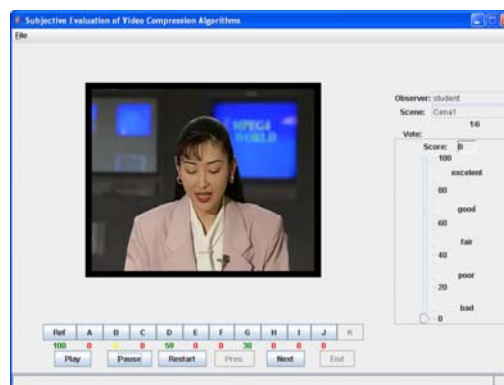


Fig. 3. SAMVIQ application.

Each session started with a brief session explaining the objectives of the test and a short demonstration. The training session is an integral part of the SAMVIQ methodology (Fig. 3). It is essential to instruct the subjects (assessors) in a special training session in advance of the tests proper. The suitable training helps to obtain more consistent results.

Videos used in the session training were not the same videos used in the actual experiment. Without the exception of the reference video, the remaining videos order was generated randomly. Viewers were only able to vote after viewing a complete video. They expressed their judgments by dragging a slider on a quality scale ranging from zero to one hundred. The highest quality should be marked “100” (top of the scale) and the lowest quality perceived should be marked “0” (bottom of the scale).

The quality scale was marked. In addition, the quality scale was divided into five equal sections labeled with adjectives: “Bad”, “Poor”, “Fair”, “Good”, and “Excellent”. After the viewers vote, the position of the slider was converted into a numeric mark by linearly mapping the scale to the interval [1 100].

Experimental results using the SAMVIQ methodology may enable subjects to arrive at more appropriate quality ratings for content that they find difficult to judge on a single viewing.

3. RATE MODEL ALGORITHM

A wide array of joint video coding/statistical multiplexing schemes have thus been developed, based primarily on the characteristics of MPEG-2 encoded video [17-23]. The results have shown this is an efficient method to obtain a uniform picture quality among video programs, while maintaining the aggregate bit rate of the various video programs conforming to the channel capacity.

Among existing solutions for joint video coding, the rate control of MPEG-2 Test Model 5 (TM5) have been the starting point for most of the proposals in the literature [17-19, 23]. The TM5 rate algorithm consists of three main steps: target bit allocation, virtual buffer based bit rate control, and adaptive quantization. However, TM5 rate control algorithm, while a good starting point, was designed for MPEG-2, a very different codec than H.264/MPEG-4 AVC. The extensive adoption of the new H.264/AVC video codec standard makes it necessary to study the video encoder's statistical characteristics and compression performance. To best of our knowledge the number of publications on this topic is still very limited [19-22]. In addition, a thorough study regarding the application of this methods regarding H.264, including objective and subjective video quality assessment, appears to be missing. We will now describe our experiences.

The Rate Control can operate at various levels of video compression, namely, sequence-level, framelevel, and macroblock (MB)-level. As MPEG does not specify how to control the bit rate several solutions have been presented in the literature. There are two different approaches: “feed forward bit rate control” and “feed backward bit rate control”.

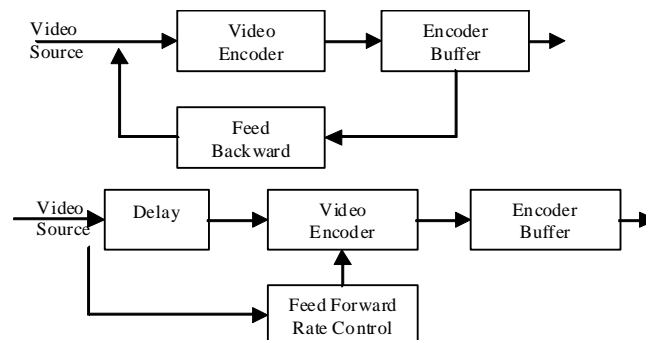


Fig. 4. Block diagram for feed-forward and feed-backward bit-rate control.

Fig. 4 presents diagrams blocks of the two different approaches. In the “feed backward bit rate control” we have limited knowledge of the sequence complexity. Statistical information is gathered by the encoders during the encoding process. This information can be used to determine the video complexity of the program. Bits are allocated on a picture basis and spatially uniform distributed throughout the image. In the “feed forward bit rate control”, a pre-analysis is performed in order to determine the optimum setting, which will increase the accuracy of the complexity metrics. In this work we have followed a feed forward bit rate control. A key decision is what statistics should be used to describe the video complexity.

Regarding statistical multiplexing for digital TV broadcast, sequence-level R-D control and optimization is performed to dynamically allocate the total bandwidth among the TV programs to maximize the statistical multiplexing gain, as well to maximize the objective quality according to the rate-distortion characteristics of the video objects [23, 24]. Thus the optimal bit allocation aims to distribute the available bit budget amongst the different programs such that the overall distortion can be minimized:

$$\min \left(\sum_{p=1}^m D_p \right) \text{ subject to } \sum_{p=1}^m R_p \leq R_{\text{target}} \quad (6)$$

where D_p and R_p denote the distortion and bit rate of the p th program respectively, m is the number of video programs. Fig. 5 presents the block diagram of the proposed approach:

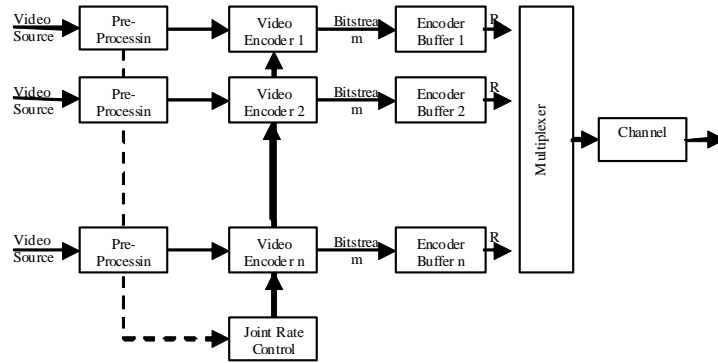


Fig. 5. Block diagram for joint coding of H.264 video programs.

Each video encoder generates video compresses bitstreams and statistics. The joint rate coding module receives information regarding the relative complexities of each program and the channel buffer fullness. Each encoder changes its bit rate only when a new GOP starts. In a practical implementation it will often be possible to send the messages from the encoder to the multiplexer in with the coded video data on the coded data link. This has two advantages: The first is that the rate control link is now reduced to an unidirectional low data-rate link. The second is that there is now no possibility of skew (unequal delay) occurring between time-critical messages generated by the encoder and the encoded video data. Thus this system will work well even if the encoder is physically remote from the multiplexer. Fig. 6 shows the required connections between a video encoder and a multiplexer in order to support statistical multiplexing.

In addition to the usual link sending coded video data from the encoder to the multiplexer, a bi-directional “rate control link” is necessary. Prior to a change in video bit-rate, a short exchange of messages is held over this link. A number of conversations are possible depending on whether the coder is asking for a modification in bit-rate from the multiplexer or the multiplexer is imposing a change of bit-rate on the encoder. Normally, the conversation proceeds as follows:

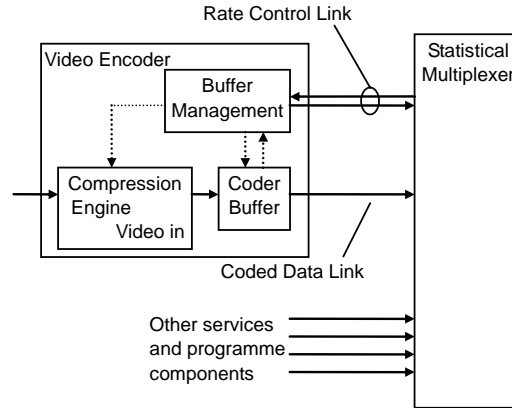


Fig. 6. Connection of a coder to a statistical multiplexer.

1. The encoder sends a message to the multiplexer requesting a new bit-rate value.
2. The multiplexer replies confirming that if it can support the new bit-rate value.
3. The encoder sends a message to the multiplexer to proceed with the adaptation of minimum and maximum encoder buffer occupancy limits.
4. On receipt of the previous message, the multiplexer waits for the codec delay and switch to the new bit-rate.

The multiplexer is the “master” device and encoders are “slaves”. The multiplexer may at any time instruct a coder to change its bit-rate and the coder must comply. Conversely, coders may request a change of bit-rate from the multiplexer but the multiplexer is under no obligation to comply. In choosing whether to satisfy an encoder’s request for more bit-rate, the multiplexer must consider many factors including bit-rate limits and service priorities assigned by the multiplex operator and requests for multiplex capacity received from other coders. These variables can change depending on the application scenario.

3.1 Algorithm

Each video encoder generates video compresses bitstreams and statistics. The joint rate coding module receives information regarding the relative complexities of each program and the channel buffer fullness. Each encoder changes its bit rate only when a new GOP begins. We assume that all the different video programs are encoded with similar GOP structures.

Let $N_{gop,p}$ designate the total number of frames in a group of picture (GOP) for the p th video program, $n_{i,j,p}$ ($i = 1, 2, \dots, j = 1, 2, \dots, N_{gop}, p = 1, 2, \dots, M$) refer to the j th frame in the i th GOP of the p th video program, $B_c(n_{i,j,p})$ represent the occupancy of virtual buffer of the p th video program after encoding the j th frame in the i th GOP, $Bit-Rate(n_{i,0,p})$ is the bitrate for encoding the i th GOP of the p th video program, and $Frame-Rate$ is the predefined frame rate.

When the encoder starts to encode the i th GOP, the total bits allocated for the i th GOP of the p th video program can be determine by the following expression:

$$T_r(n_{i,0,p}) = \frac{\text{BitRate}(n_{i,0,p})}{\text{FrameRate}} N_{\text{GOP}} - B_c(n_{i-1,N_{\text{GOP}},p}) \mid_{p=1,2,\dots,M}. \quad (7)$$

If we consider the independent case this would be the available bandwidth. The aggregate bandwidth of all the video programs for the i th GOP can be determined as follows:

$$T_{\text{aggregate}_i} = \sum_{p=1}^M T_r(n_{i,0,p}). \quad (8)$$

In the case of joint video coding, bandwidth is allocated proportional to its ratio complexity. Thus the bandwidth for the i th GOP of the p th video program will be compute as:

$$T_{i,p} = \frac{X_{i,p}}{\sum_{t=1}^M X_{i,t}} \times T_{\text{aggregate}_i} \quad (9)$$

where $X_{i,p}$ is the complexity of the program p for the i th GOP and M the number of video program that are jointly encode. In our design we use SSIM metric to measure the distortion of the encoded sequences.

3.2 Results

We have implemented the proposed rate control scheme using the H.264 JM 10.2 encoder [26]. Simulations were performed using several video test sequences (Table 1), representing different level of complexity, each one exhibiting different combined levels of spatial detail and amount of movement.

Table 1. Test sequences.

Name	Res.	Duration	Characteristics
Akiyo	CIF	10 sec.	Still camera on human subject with i th synthetic background
Foreman	CIF	10 sec.	Fast camera and content motion with pan at the end
Football	CIF	10 sec.	Fast translational motion and camera panning; moderate spatial detail

The test conditions under which our experiments were conducted are as follow: MV resolution = 1/4 pel; RDO = ON; Search Range = 32; Frame Type = IPPP and Reference Frame = 1. The performance of our proposed scheme is evaluated in comparison with the original encoder JM 10.2. Two scenarios were studied. First each sequence encoded at fixed bit rate of 256kbps and then at fixed 512kbps. Table 2 presents results for independent coding (CBR). GOP1 and GOP2 correspond to Intra Period of 4 and 10 frames respectively.

With these three sequences 3 testing groups of two video sources (Table 3) were created. Several combinations were simulated with jointly combine 3 video streams: Akiyo is represented with letter A, Foreman with letter B and Football with letter C.

Table 2. PSNR for independent coding (dB).

Sequence	GOP	PSNRY (256kbps)	PSNRY (512kbps)
Akiyo	IPPP_GOP1	37,90	41,82
	IPPP_GOP2	41,21	44,06
Foreman	IPPP_GOP1	31,13	34,40
	IPPP_GOP2	33,05	36,18
Football	IPPP_GOP1	28,52	31,90
	IPPP_GOP2	29,18	31,91

Table 3. Composition of group of two video programs.

Group Name	Sequence Name	Sequence Name
AB	Akiyo	Foreman
AC	Akiyo	Football
BC	Foreman	Football

Table 4. Statistical multiplexing with SSIM.

Sequence	BitRate	Group	IPPP_GOP1	IPPP_GOP2
Akiyo	256	AB	36,04	36,43
Foreman		AC	34,31	34,13
		AB	32,95	35,33
		BC	30,23	32,11
		Football	AC	30,55
BC			29,49	30,29
Akiyo	512	AB	39,95	40,07
Foreman		AC	38,57	37,23
		AB	36,25	38,25
		BC	33,63	35,30
		Football	AC	33,50
BC			32,31	33,13



(a)



(b)

Fig. 7. Akiyo (frame 35), AAC, 256 kbps (a) encoded JM10.2, (b) statistical multiplexing with SSIM.



Fig. 8. Football (frame 35), AAC, 256 kbps (a) encoded JM10.2, (b) statistical multiplexing with SSIM.

This paper presents an algorithm for dynamic bandwidth allocation which allots the available bandwidth according to the needs of each video source and perceptual importance. Video sources with different combined levels of spatial detail and amount of movement benefit from this method. The main characteristics of SSIM index are the use of structural distortion as an alternative of error sensitivity and its simplicity. SSIM implementation does not require complex procedures (such as spatial and temporal filtering, object segmentation, texture classification, or blockiness estimation). Simulation results show that bandwidth gains/quality improvements are more significant when heterogeneous sources are multiplexed together. An increased in subjective quality may be observed in football sequence while the decrease in Akiyo is rather small.

REFERENCES

1. ITU-T and ISO/IEC JTC 1, "Advanced video coding for generic audiovisual services," ITU-T Rec. H.264 & ISO/IEC 14496-10, Version 1, 2003; Version 2, 2004; Version 3 (with High family of profiles), 2004; Version 4, 2005.
2. T. Wiegand and G. J. Sullivan, "The H.264/AVC video coding standard [standards in a nutshell]," *IEEE Signal Processing Magazine*, Vol. 24, 2007, pp. 148-153.
3. T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, 2003, pp. 688-703.
4. K. Seshadrinathan and A. C. Bovik, "New vistas in image and video quality assessment," in *Proceedings of SPIE Human Vision and Electronic Imaging*, Vol. 6492, 2007, pp. 649202.1-649202.13.
5. H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, Vol. 15, 2006, pp. 430-444n.
6. W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error measurement to structural similarity," *IEEE Transactions on Image Processing*, Vol. 13, 2004, pp. 600-613.
7. H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, Vol. 15, 2006, pp. 3440-3451.
8. O. Nemethova, M. Ries, E. Siffel, and M. Rupp, "Subjective evaluation of video

- quality for H.264 encoded sequences,” in *Proceedings of Joint IST Workshop on Mobile Future and Symposium on Trends in Communications*, 2004, pp. 191-194.
9. T. Wolff, H. H. Ho, J. M. Foley, and S. K. Mitra, “H.264 coding artifacts and their relation to perceived annoyance,” in *Proceedings of European Signal Processing Conference*, 2006.
 10. M. Farias, “No-reference and reduced reference video quality metrics: New contributions,” Ph.D. Thesis, Department of Electrical and Computer Engineering, University of California, 2004.
 11. A. C. Brooks and T. N. Pappas, “Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions,” in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2007, pp. 869-872.
 12. C. H. Chou and C. W. Chen, “A perceptually optimized 3-D subband image codec for video communication over wireless channels,” *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 6, 1996, pp. 143-156.
 13. Z. Wang, A. C. Bovik, and E. P. Simancelli, “Structural approaches to image quality assessment,” *The Handbook of Image and Video Processing*, 2nd ed., Chapter 8.3, Academic Press, 2005
 14. Z. Wang and X. Shang, “Spatial pooling strategies for perceptual image quality assessment,” in *Proceedings of IEEE International Conference on Image Processing*, 2006, pp. 2945-2948.
 15. EBU BPN 056: SAMVIQ – Subjective Assessment Methodology for Video Quality Report by EBU Project Group B/VIM (Video In Multimedia), 2003.
 16. F. Kozamernik, V. Steinman, P. Sunna, and E. Wyckens, “SAMVIQ – A new EBU methodology for video quality evaluations in multimedia,” *SMPTE Motion Imaging Journal*, Vol. 114, 2005, pp. 152-160.
 17. L. Wang and A. Vincent, “Joint rate control for multi-program video coding,” *IEEE Transactions on Consumer Electronic*, Vol. 42, 1996, pp. 300-305.
 18. L. Boroczky, A. Y. Ngai, and E. F. Westermann. “Joint rate control with look-ahead for multi-program video coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 10, 2000, pp. 1159-1163.
 19. Z. G. Li, C. Zhu, F. Pan, G. Feng, X. Yang, S. Wu, and N. Ling, “A novel joint rate control scheme for the coding of multiple real time video programs,” in *Proceedings of the 22nd International Conference on Distributed Computing Systems*, 2002, pp. 241-245.
 20. S. Kwon, K. R. Rao, Q. J. Kwon, and T. S. Kim, “Joint bandwidth allocation for user-required picture quality ratio among multiple video sources,” *IEEE Transactions on Broadcasting*, Vol. 51, 2005, pp. 287-295.
 21. L. Teixeira and L. Corte-Real, “Statistical multiplexing of H.264 programs,” in *Proceedings of International Conference on Intelligent and Advanced Systems*, 2007, pp. 621-625.
 22. J. Yang, X. Fang, and H. Xiong, “A joint rate control scheme for H.264 encoding of multiple video sequences,” *IEEE Transactions on Consumer Electronics*, Vol. 51, 2005, pp. 617-623.
 23. M. Perkins and D. Arnstein, “Statistical multiplexing of multiple MPEG-2 video programs in a single channel,” *SMPTE Journal*, Vol. 104, 1995, pp. 596-599.
 24. L. Teixeira and T. Andrade, “Exploiting characteristics of a large number of MPEG

video sources for statistically multiplexing video for TV broadcast applications,” in *Proceedings of the 4th Bayona Workshop on Intelligent Methods in Signal Processing and Communications*, 1996, pp. 65-69.

25. Z. Chen and K. N. Ngan, “Optimal bit allocation for MPEG-4 multiple video objects,” in *Proceedings of IEEE International Conference on Image Processing*, Vol. 2, 2004, pp. 761-764.
26. Joint Video Team (JVT), “H.264/Advanced video coding reference software version 10.2,” 2006, <http://iphome.hhi.de/suehring/ttml/>.



Luís Teixeira (M'95) was born in Coimbra, Portugal, in 1969. He received the degree in Electrical Engineering from the Faculdade de Engenharia, Universidade do Porto, Porto, Portugal, in 1992, the M.Sc. degree in Electrical and Computer Engineering Faculdade de Engenharia, Universidade do Porto, Porto, Portugal, in 1995, and he is currently pursuing the Ph.D. degree at Departamento de Engenharia Electrotécnica e de Computadores da Faculdade de Engenharia da Universidade do Porto. Since 1992, he has been a Researcher at INESC PORTO, an R&D institute affiliated with Universidade do Porto. In 1995, he joined the Universidade Catolica Portuguesa as a Lecture at Arts School. His research interests include video coding, video quality assessment, and perceptual video analysis.



Luís Corte-Real (M'91) was born in Vila do Conde, Portugal, in 1958. He received the degree in Electrical Engineering from the Faculdade de Engenharia, Universidade do Porto, Porto, Portugal, in 1981, the M.Sc. degree in Electrical and Computer Engineering from the Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisbon, Portugal, in 1986, and the Ph.D. degree from the Faculdade de Engenharia, Universidade do Porto, in 1994. In 1984, he joined the Universidade do Porto as a Lecturer of telecommunications. He is currently an Associate Professor with the Departamento de Engenharia Electrotécnica e de Computadores da Faculdade de Engenharia da Universidade do Porto. Since 1985, he has been a Researcher at INESC Porto, an R&D institute affiliated with Universidade do Porto. His research interests include image/video coding and processing and content-based image/video retrieval.